To: Tina Laidlaw/MO/R8/USEPA/US@EPA;Dave Moon/R8/USEPA/US@EPA[]; ave Moon/R8/USEPA/US@EPA[]
Cc: "Urban, Eric" [EUrban@mt.gov]
From: "Suplee, Mike"
Sent: Wed 6/27/2012 4:26:25 PM
Subject: RE: Nutrient Standards M/D/F and Assessment Methods
Memo_StndtoAssessMethod_Compared_FNL.pdf
http://deq.mt.gov/wqinfo/qaprogram/sops.mcpx

Hi Tina, Dave;

The attached memo is our analysis and response to your concerns that the proposed duration and frequency aspects of the base numeric nutrients may not be in alignment with the Department's finalized nutrient assessment methodology for streams. The nutrient assessment methodology is located at:

http://deq.mt.gov/wqinfo/qaprogram/sops.mcpx

It is shown as "Nutrient Assessment Method" near the middle of the webpage.

If you have any questions regarding the analysis or the results in the memo, please do not hesitate to contact me by E-mail or phone (406) 444-0831.

Thanks,

Mike

Montana Department of
ENVIRONMENTAL QUALITY

MEMO

1520 E 6th Avenue
PO Box 200901
Helena MT 59620-0901
(406) 444-6697 □ FAX: (406) 444-3836

To: Dave Moon, Tina Laidlaw, Region VIII Environmental Protection Agency

CC: Eric Urban, Supervisor, Water Quality Standards Section

From: Michael Suplee, Ph.D., Water Quality Standards Section

Date: 6/27/2012

RE: Linkage between the Department's Eutrophication Assessment Methodology and Magnitude, Duration, and Frequency of the Base Numeric Nutrient Standards

The Department will be proposing to the Board of Environmental Review that they consider numeric nutrient criteria for adoption as water quality standards. The purpose of this memo is to clarify the linkage between statistical methods used in the Department's stream eutrophication assessment methodology and key elements of the new standards, specifically their magnitude, duration, and frequency.

Magnitude. Base numeric nutrient standards will be housed in a new Department circular, Circular DEQ-12 Part A. Derivation of the criteria magnitudes is detailed in Suplee et al. (2008) and Suplee and Watson (2012). The standards are expressed as water concentrations (µg/L) for total nitrogen (TN) and total phosphorus (TP) and have a period of application of (usually) July through September of each year. In Section 3.2.1 of the Department's eutrophication assessment methodology (Suplee and Sada de Suplee, 2011), readers are instructed to refer to Circular DEQ-12 to locate the TN and TP concentration values they are to use.

Duration. Base numeric nutrient standards will have a duration of 30 days as the averaging period. Studies by the Department and others show that when nutrient concentrations are elevated it takes about 15-20 days for benthic algal biomass to develop to the point where it becomes a nuisance for recreation (Stevenson et al., 1996; Flynn and Suplee, 2011; Suplee and Sada de Suplee, 2011) and, if the stream is wadeable, this may also lead to violations of the dissolved oxygen standards later in the fall when the algae senesce and decompose *en masse* (Appendix B, Suplee and Sada de Suplee, 2011). A primary intent of the lotic numeric nutrient standards is to control excess benthic algae and, as such, the growth patterns of benthic algae were used to derive the acceptable averaging period. The Department is confident that if numeric nutrient standards were exceeded (for example) for only a few consecutive days in a month, it would be an insufficient amount of time for algae to develop to problem levels.

Frequency. For the base numeric nutrient standards the Department is proposing that "the 30 day (monthly) average concentration of these parameters may not be exceeded more than once in any five year period, on average". Thus, the frequency of allowable exceedence is, for all practical purposes, one year (one summer) in 5, or about 20% of the years, which is in keeping with EPA recommendations for aquatic life protection (USEPA, 1985).

Linkage between the Numeric Nutrient Standards and the Department's Assessment Methodology. As outlined above, the nutrient standards will have a 30 day averaging period (duration) and, for all practical purposes, a one summer in every five years (1 in 5) allowable exceedence frequency. The Department's eutrophication assessment methodology (Suplee and Sada de Suplee, 2011) comprises statistical evaluation of nutrient-concentration datasets as well as consideration of biological data (e.g., biometrics based on macroinvertebrate taxa). Here, however, the focus will be on just one of the statistical tools used in the methodology to evaluate nutrient concentration datasets—the exact binomial test. In the assessment methodology, the exact binomial test is used to determine whether a nutrient dataset from a stream reach demonstrates a >20% exceedence rate for either a TP or TN criterion (i.e., > 1 in 5). This allowable exceedence rate was derived empirically from an analysis of twelve years of data at ten sites along the Clark Fork River, where the state has had numeric nutrient standards in place since 2002. The analysis showed that, on average, when the river's nutrient standards were exceeded at a site more than about 25% of the time during a summer, algal biomass would exceed the nuisance threshold (150 mg Chl*a*/m^2). In contrast, when a site manifested lower exceedence rates (< 20%) nuisance algae levels rarely develop. As a result, the Department established a 20% exceedence threshold for the exact binomial test when evaluating other streams' nutrient datasets per the eutrophication assessment methodology.

To explore whether the assessment methodology will be consistent with the proposed magnitude/ duration/frequency of the nutrient standards, a real-world dataset was examined. The Clark Fork River dataset was used, as it is among the Department's most complete long-term ambient monitoring records. Some sites were sampled on up to three different days per month. This frequency of monthly monitoring (3/month) was adequate to be able to compare decisions that would be made via the monthly averaging approach vs. the exact binomial test approach, the former of which would be a direct interpretation of the duration component of the new standards. If the two methods provide comparable results, i.e., if (at a site) there is a similar number of months found to exceed the standard and a similar number of months that are in compliance, it would suggest that the exact binomial test as used in the assessment methodology is largely in alignment with the way the nutrient standards will be expressed in Circular DEQ-12.

For each Clark Fork River monthly dataset, the number of exceedences was determined by comparing the monthly averages to the applicable criterion magnitude. This met the direct intent of the standards. Then, the number of monthly criteria exceedences identified via the binomial test was calculated; the binomial test was programmed with the same parameterizations (alpha error, effect size, etc.) established in the assessment methodology. The two methods provided fairly similar results, although the average-monthly approach consistently identified more exceedences (Table 1). The monthly-average approach resulted in 9% (on average) more monthly exceedences of the criteria than did the exact binomial test.

Table 1. Exceedence Frequencies Compared Between the Monthly Average Approach and the Exact Binomial Test Using Clark Fork River Data.

							% Non-compliant Months	
Site Name	1998-2009: Number of months with 3 sampling events per month	Nutrient	Criterion (μg/L)	Compliance Threshold in Exact Binomial Test (%)	Months Found Non-compliant, Monthly Average	Months Found Non-compliant, Exact Binomial	Monthly Average	Exact Binomial
Clark Fork above Little Blackfoot (10)	23	P	20	20%	20	18	87%	78%
Clark Fork above Little Blackfoot (10)	23	N	300	20%	10	8	43%	35%
Clark Fork above Missoula (15.5)	9	N	300	20%	1	0	11%	0%

Clark Fork River site 10 has a long history of elevated nutrients and algae problems and site 15.5 does not, and the monthly-average approach and the binomial approach both reflect this fact. The 9% difference between the two test approaches is likely due to low sample size (n = 3 samples/month). For the exact binomial test, at this low sample size the beta error (false negative rate, i.e., the probability that a truly non-compliant month is declared as compliant) is quite high (72% chance). Thus, the additional one to two months found to be in compliance by the binomial test are likely the result of beta error.

For the assessment methodology the Department generally requires 12 TN or TP samples for routine stream evaluation. To test how the monthly average and the binomial methods would compare if twelve samples collected during a month were actually available, 200 TN datasets of n=12 each were randomly generated using a bootstrap program. The bootstrap program randomly re-samples (with replacement) the source dataset and creates a new one; this process may be repeated as much as desired (200 cases should be more than adequate to elucidate the overall pattern). The source dataset contained individual TN samples equal to, above, and below 300 µg TN/L[1] (the assumed criterion), and had an average of 320 µg TN/L (Table 2). In applying the two test methods to the 200 randomly-generated datasets, it was found that the average-monthly approach indicated 59% of the months exceeded the criterion and the exact binomial test indicated 71% of the months exceeded the criterion (Table 2). Interestingly, these larger (albeit artificial) datasets indicate that the binomial test was, on average, about 12% *more* protective than the monthly-average approach. (This contrasts with the results in Table 1 because, as discussed, the low monthly samples size of the Clark Fork datasets causes high false-negative rates.) A sensitivity analysis of the same generated datasets further showed that the binomial test will move to a less protective stance (43% of months found to exceed) when the binomial compliance threshold is increased to just 22% (Table 2). The gap of about 14% on each side of the monthly average results manifested by the binomial test (Table 2, last two columns) occurs due to the Department having selected a 15% effect size for the binomial test.

Table 2. Exceedence Frequencies Compared Between the Monthly Average Approach and the Exact Binomial Test Using Generated Data.

							% Non-compliant Months	
Number of Generated Monthly Datasets with 12 samples/month	Nutrient	Overall Mean of Generated Datasets (µg TN/L)	Criterion (µg/L)	Compliance Threshold in Exact Binomial Test (%)	Months Found Non-compliant, Monthly Average	Months Found Non-compliant, Exact Binomial	Monthly Average	Exact Binomial
200	N	320	300	20%	118	142	59%	71%
200	N	320	300	22%	118	85	59%	43%

Conclusion. The proposed nutrient standards will have a 30-day averaging period and a 1 in 5 year allowable exceedence rate. The established compliance threshold of 20% (1 in 5) used in the exact binomial

[1] Concentrations in the source dataset ranged from 50 to 900 µg TN/L.

test of the Department's eutrophication assessment methodology is reflective of the standard's frequency. Further, the binomial test—when used in cases where the required 12 samples have been collected— leans slightly to the protective side of the results observed using the monthly-averaging approach. As such, the Department believes that the current assessment methodology should be largely in alignment with the magnitude, duration, and frequency aspects of the proposed numeric nutrient standards.

REFERENCES

Flynn, K. and M. W. Suplee, 2011. Using a Computer Water Quality Model to Derive Numeric Nutrient Criteria: Lower Yellowstone River. WQPBDMSTECH-22. Helena, MT: Montana Dept. of Environmental Quality

Stevenson, R. J., M. L. Bothwell, and R. Lowe, 1996. Algal Ecology - Freshwater Benthic Ecosystems. San Diego, CA: Acaemic Press.

Suplee, M.W., V. Watson, A. Varghese, and J. Cleland, 2008. Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers. Helena, MT: Montana Department of Environmental Quality, 86 p.

Suplee, M.W., and V. Watson, 2012. Addendum 1: Scientific and Technical Basis of the Numeric Nutrient Criteria for Montana's Wadeable Streams and Rivers. Final Draft. Helena, MT: Montana Department of Environmental Quality.

Suplee, M.W., and R. Sada de Suplee, 2011. Assessment Methodology for Determining Wadeable Stream Impairment Due to Excess Nitrogen and Phosphorus Levels. WQPBMASTR-01, Helena, MT: Montana Dept. of Environmental Quality.

USEPA (United States Environmental Protection Agency), 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. Office of Research and Development, Environmental Research Laboratories, Duluth, MN, Narragansett, RI, and Corvallis, OR. PB85-227049.

USEPA (United States Environmental Protection Agency), 1991. Technical Support Document for Water Quality-based Toxics Control. Office of Water. EPA/505/2-90-001.